# EGI towards an Open Science Commons


## Tiziana Ferrari

EGI.eu Technical Director


www.egi.eu


EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142


# Welcome to Lisbon!

# Outline

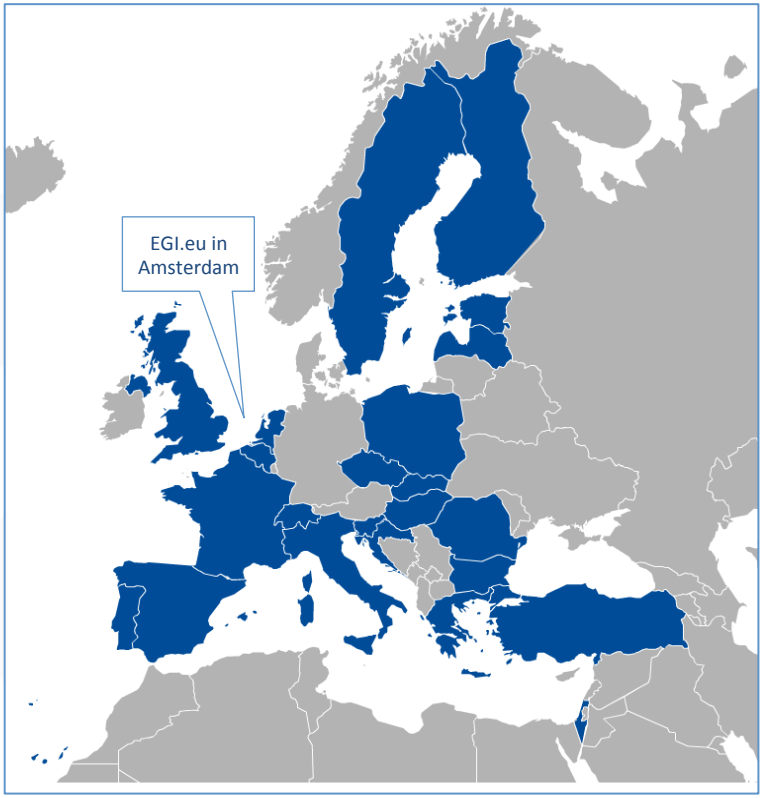
- EGI today
- Medium-term plans
- Towards 2020

# EGI today

- Governance, the power of federating

# EGI and its participants - 2015

- 25 participants: 23 NGIs and 2 EIROs (CERN, EMBL-EBI)
  - Opening membership to research communities
- Affiliation programme
  - lower barriers of entry to widening countries




**Participants**

CERN, EMBL-EBI, Belgium, Bulgaria, Croatia, Czech Republic, Estonia, Finland, France, Greece, Hungary, Israel, Italy, FYR of Macedonia, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Switzerland, Sweden, Turkey, UK

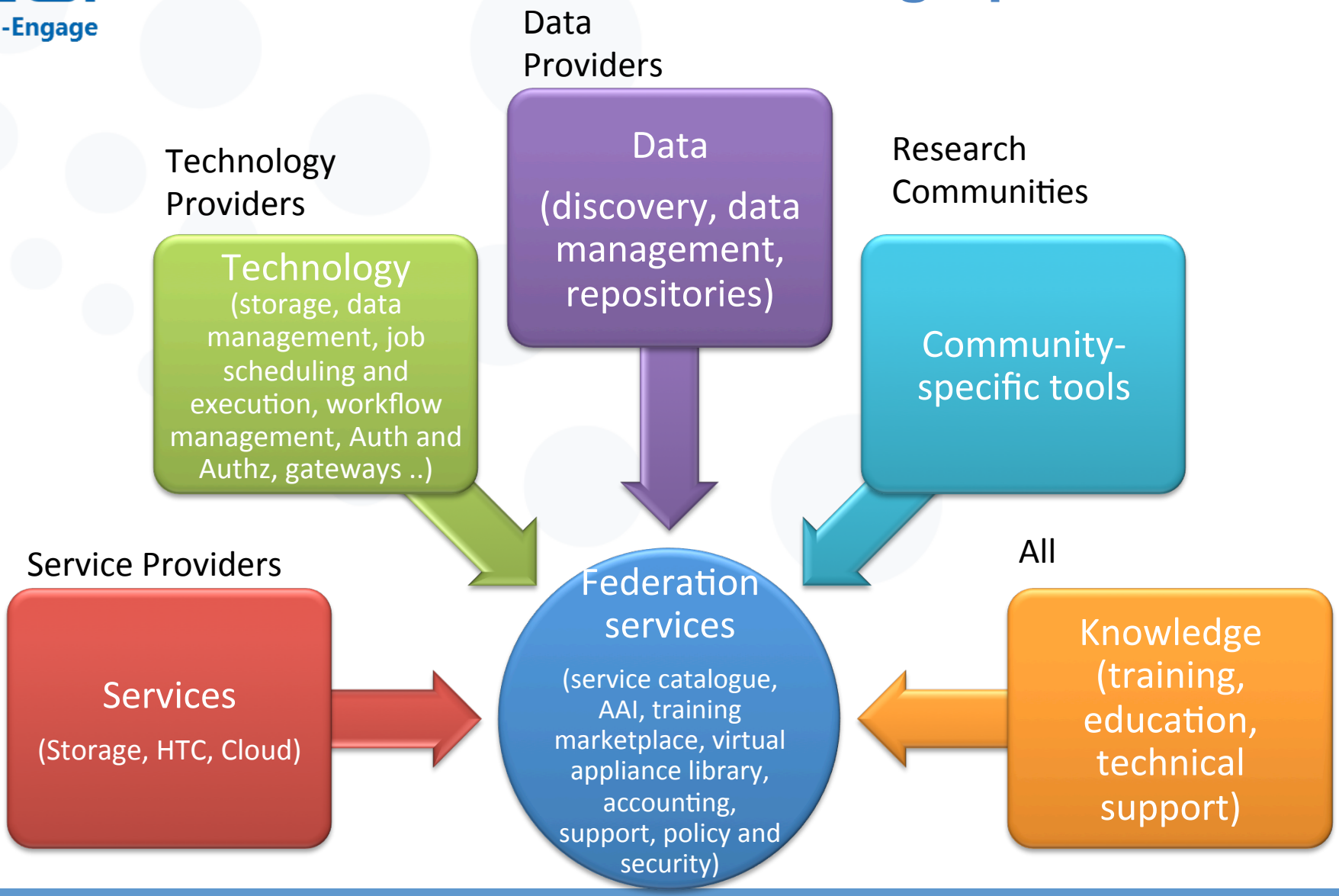
**Under discussion**

Armenia, Austria, Belarus, Germany, Denmark, Moldova, Norway, Russia, Ukraine

# EGI Offer

- High-Throughput Data analysis
- Federated Cloud
- Federated Open Data Processing
- Federated Operations
- Community driven Innovation and Support
- Policy Advice

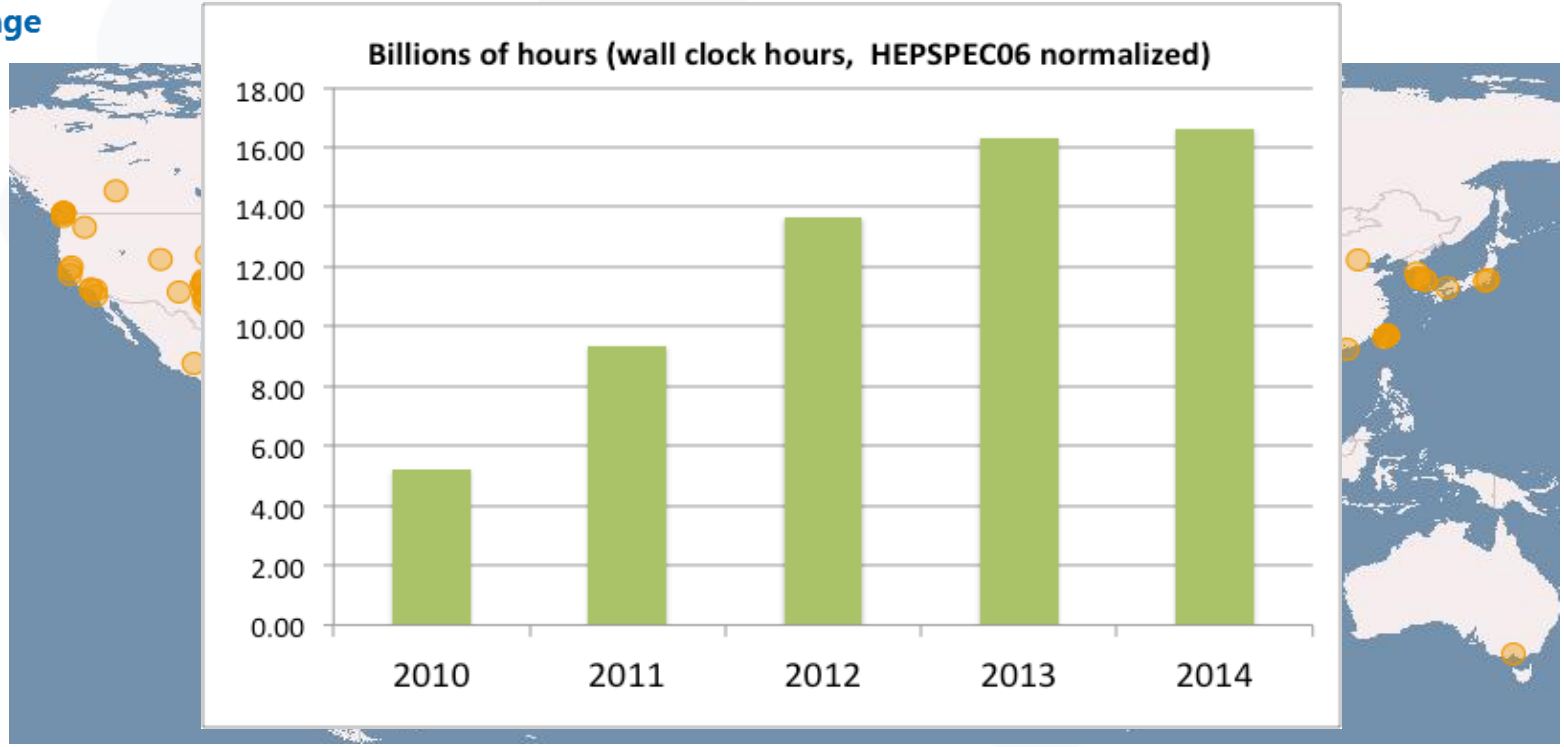
# Federating open science



# Science is inherently distributed

- Discoverability of services and knowledge
- Portability
  - data, applications, software
- Sharing and openness
- Common access policies, security
- One accounting infrastructure
- One support infrastructure
- Single sign on
- Federated service management
- Aggregation of demand and offer

# Federating e-infrastructures and data 1/2




- Distributed, federated storage, HTC and cloud facilities
- Virtual Research Environments
- > 200 registered user research projects

- 340 resource centres in 54 countries
- 550,000 logical CPU cores
- >290 PB disk, 180 PB tape
- > 99.6% reliability

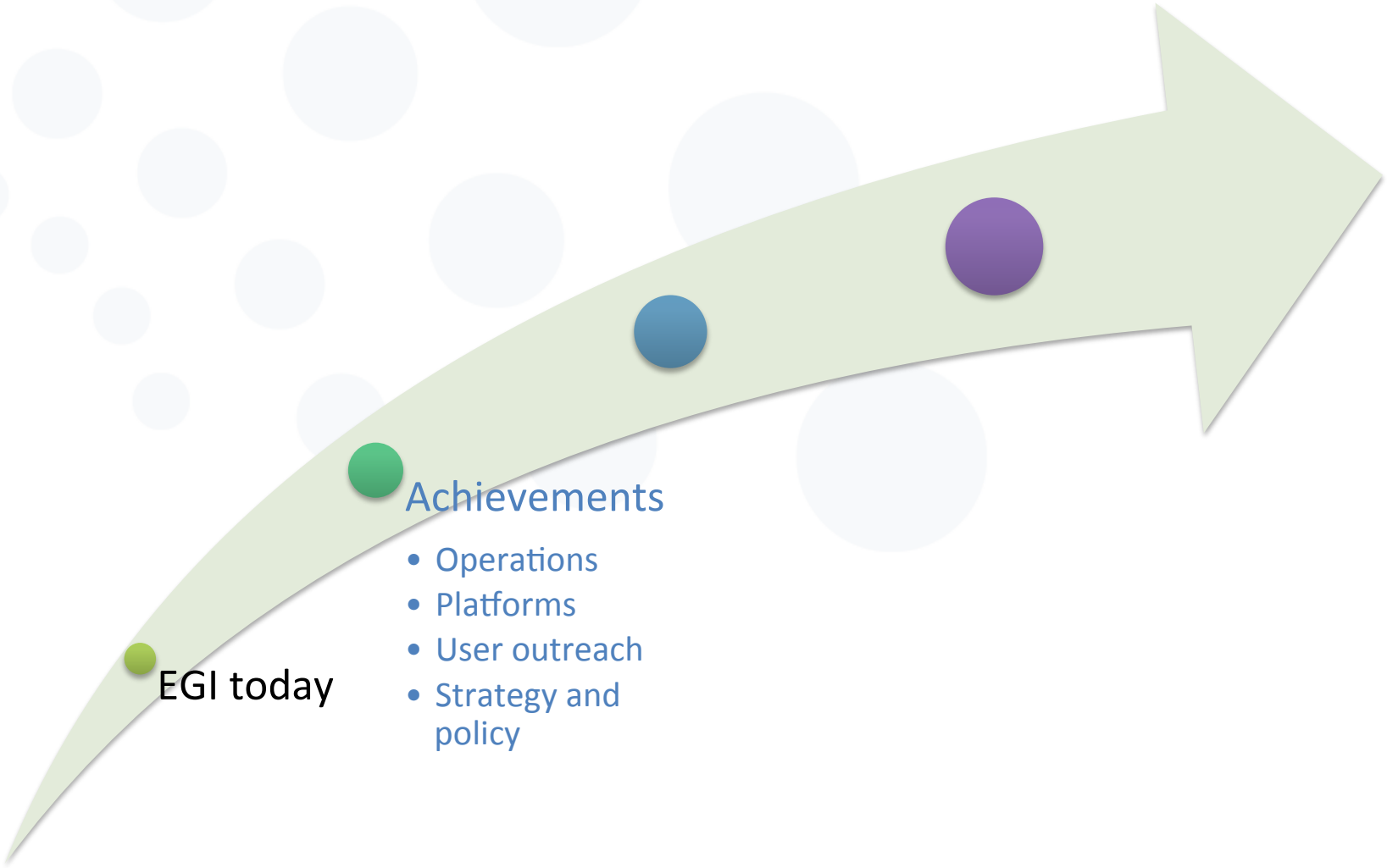
# Federating e-infrastructures and data 2/2

EGI-Engage

EGI Council members
Integrated Infrastructures
Peer Infrastructures


→ More than 6,000 jobs/year to OSG
→ More than 68,000 jobs/year in IDGF
→ 840 M CPU hours/year in Asia Pacific

d4science
iMarine
Canadian Astronomy Data Centre
E-CEO E-Collaboration For Earth Observation


Achievements
• Operations
• Platforms
• User outreach
• Strategy and policy
EGI today

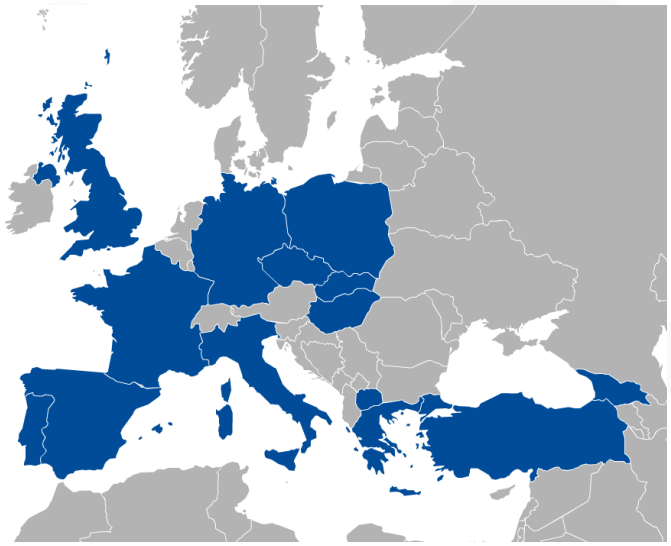
# Get infrastructure services

Resource allocation for national and international resources

- e-GRANT
  - Pooling of distributed infrastructure resources (HTC and cloud)
  - Matchmaking demand ←→ offer
  - Allocation
  - SLA negotiation (user community ←→ EGI.eu)
- Monitoring of service level targets

Wed 20/05: Service Level Management for federated e-Infrastructures

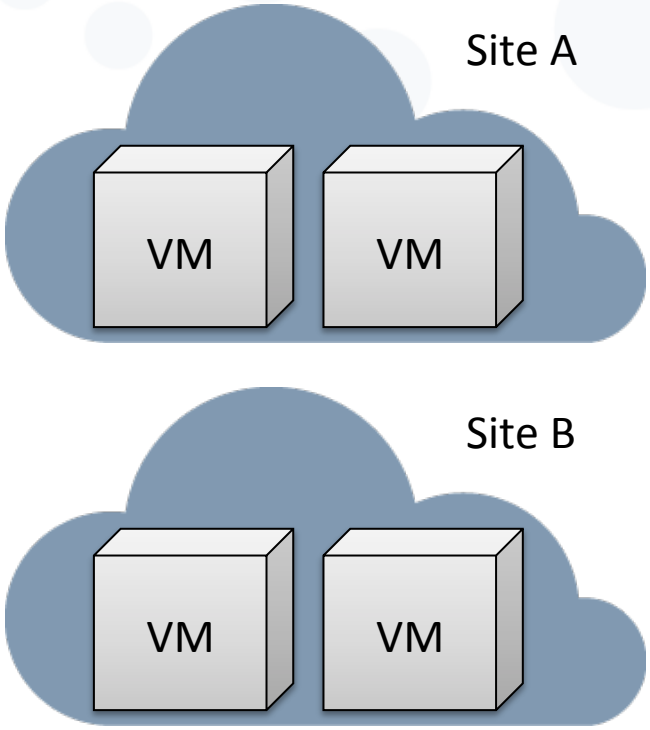
# EGI Federated cloud



- Hybrid federation
  - Public clouds (open to any research community, based on open cloud standards for portability of applications and data)
  - Community clouds (for selected list of VOs, looser federation profile based on a subset of federation tools)
  - Bringing cloud services next to big data
  - Federated AAI, accounting, discovery and monitoring

# VM Management

## On demand compute to run any kind of workloads on virtual machines




**Easy provisioning**

- OCCI API across the whole infrastructure
- VMs start immediately
- Ruby and Java clients

**Customize**

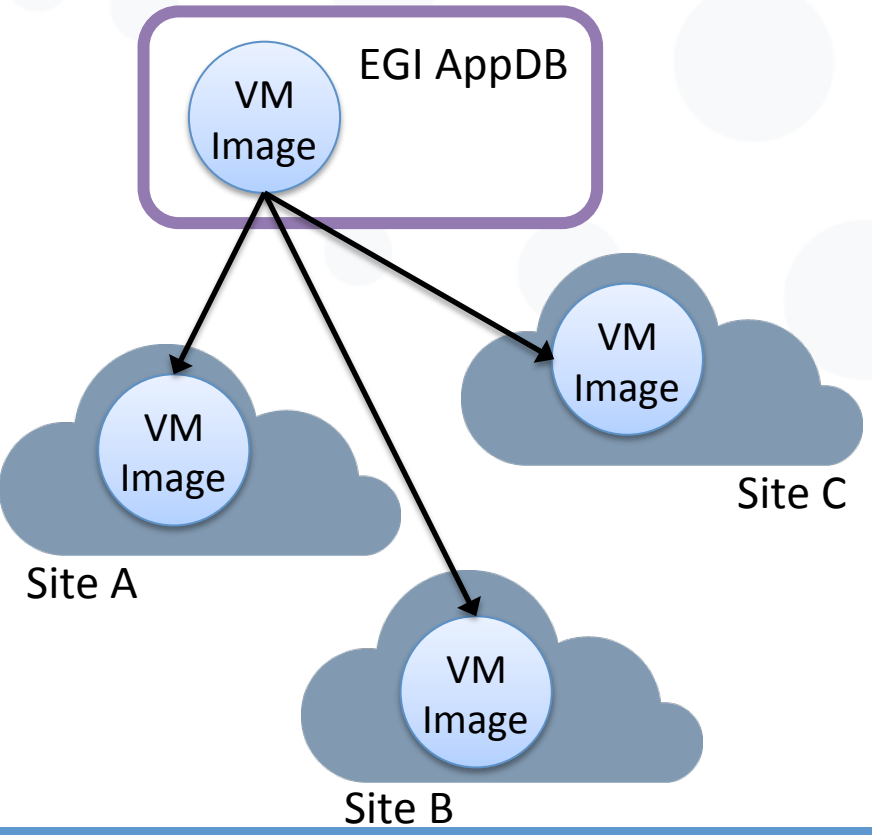
- Select your OS
- *root* access
- Contextualization

**Scale to your needs**

- Select VM size (cores, RAM)
- Create and destroy VMs as needed

# VM Image Management

## Automatic and secure distribution of endorsed VM images for Virtual Organisations




**Web based**

- Easy creation from AppDB
- Re-use and extend images

**EGI endorsed images**

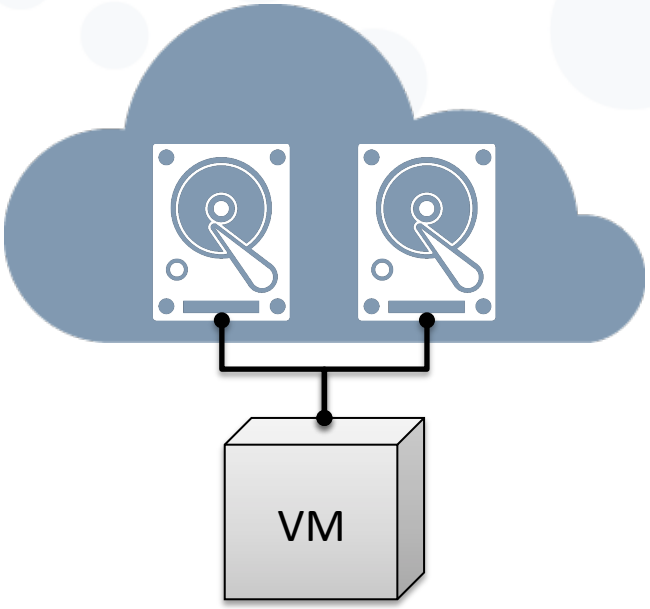
- Basic OS ready to use and contextualize
- Available on every site

**VO-level control**

- VO endorse images
- Automatically distributed to sites supporting the VO

# Block Storage

## Persistent Block Level Storage to attach to VMs




**Simple usage**

- Manage with OCCI and use as any other block device from VMs (i.e. POSIX)
- Snapshotable

**High Performance**

- Consistent and low-latency performance
- SSDs (in some sites)

**Scale to your needs**

- From GB to TB
- Create and attach to VMs on demand

# Object Storage

**Data storage infrastructure for storing and retrieving data from anywhere at any time**

| | |
|---|---|
| API Access | • CDMI REST API for managing and accessing data |
| Sharing | • Define ACLs on each object, share publicly your data |
| Scale to your needs | • Store as much data as needed<br>• Get accounted only for the space used |

# 12 months of Federated Cloud activities

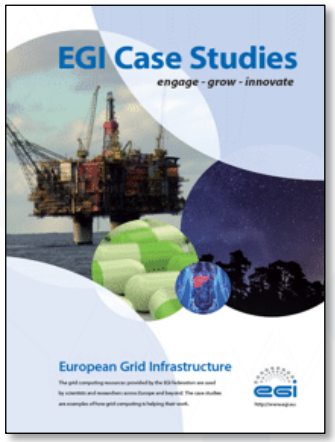
- 26 communities
  - Biological sciences
  - Physical sciences
  - Earth sciences
- 59 use cases currently supported, 5 from commercial organisations
- 700,000 VMs instantiated

➡ Tuesday session: Federated IaaS track

# Strategy, Policy, Business Development

- New EGI strategy for 2020 in consultation with the EGI.eu Executive Board and the EGI Council
- The Open Science Commons
- Pay-for-Use pilot
  - Pay 30 providers across 12 countries publishing pricing information (~10 ready/able to sell)
  - Emerging business models
  - Tools adapted (GOCDB, AppDB, e-GRANT), including GUI
  - Final Report

# Impact




[EGI case studies]

3,600 service end-points, 47 UMD releases,38,000 users

Increasing use of new disciplines

- 220 research projects, 76 new
- Astronomy and astro-particle Physics, Structural biology, Hydrology and climate, Medical and Health Sciences

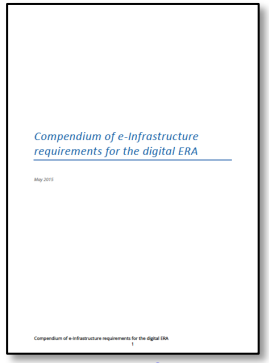
Better services for the long tail

- 46% of the new users)

Support to Research Infrastructures

- BBMRI, CTA
- Testing: EISCAT-3D, ELIXIR, ELI-NP, LifeWatch, LOFAR, KM3NeT

2,400 Peer-reviewed papers, 620 new registered applications




[Compendium of RI requirements]


EGI today
Achievements
The big shifts through EGI-Engage and sister projects
• AARC
• INDIGO Datacloud
• EDISON
• Technical support (ENVRI Plus, ELITRANS)

# The big shifts

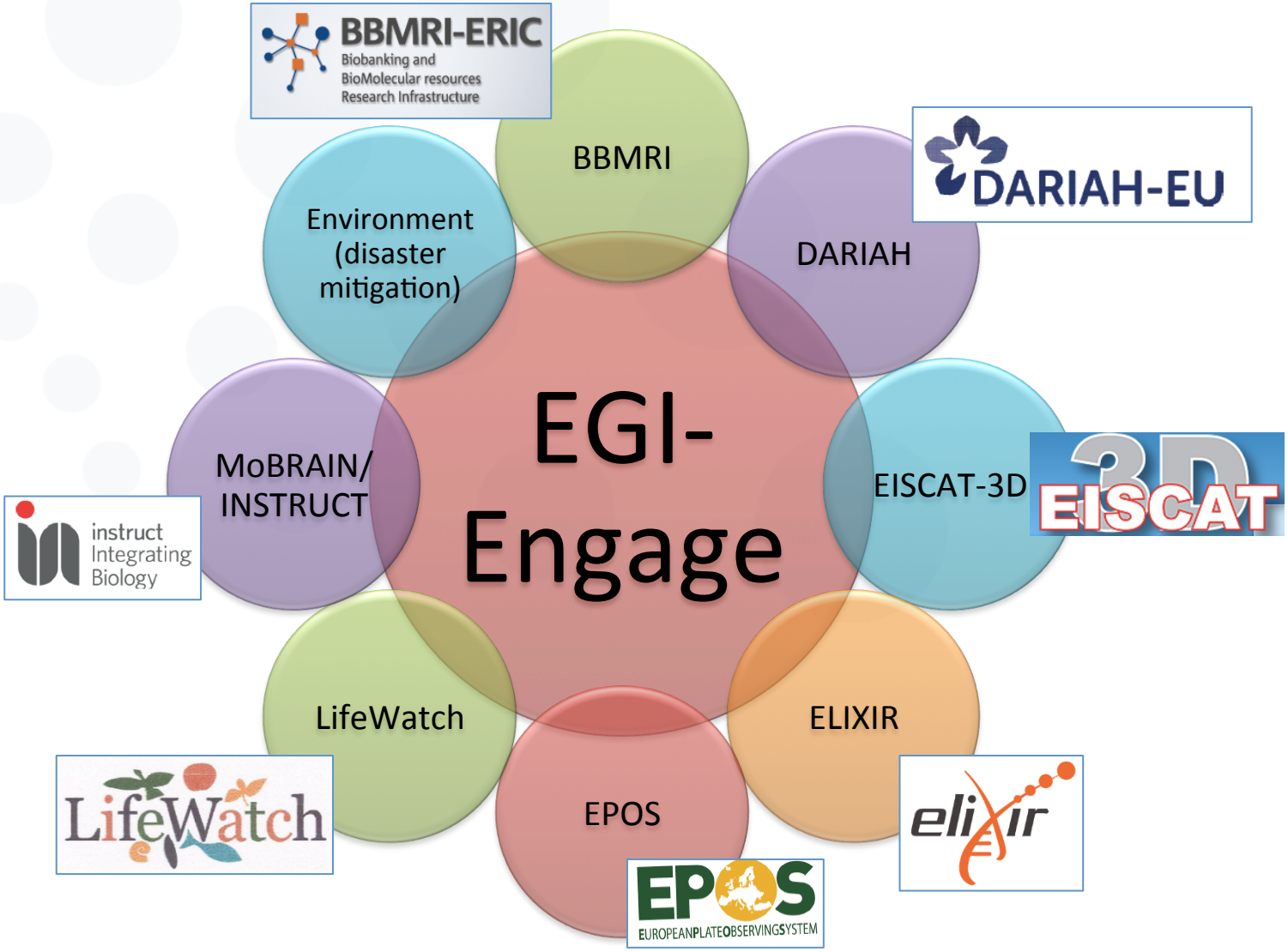
New governance to community engagement

The Distributed Competence Centre

# Distributed Competence Centre (DCC)

- Promote reuse of solutions of common interest across research communities
- Evolve the EGI technical services with community requirements and provide a test environment with NGIs/EIROs → co-development
- Promote the integration of community services
  - Scientific applications
  - Joint training programme
  - Technical user support

# EGI-Engage support to the DCC

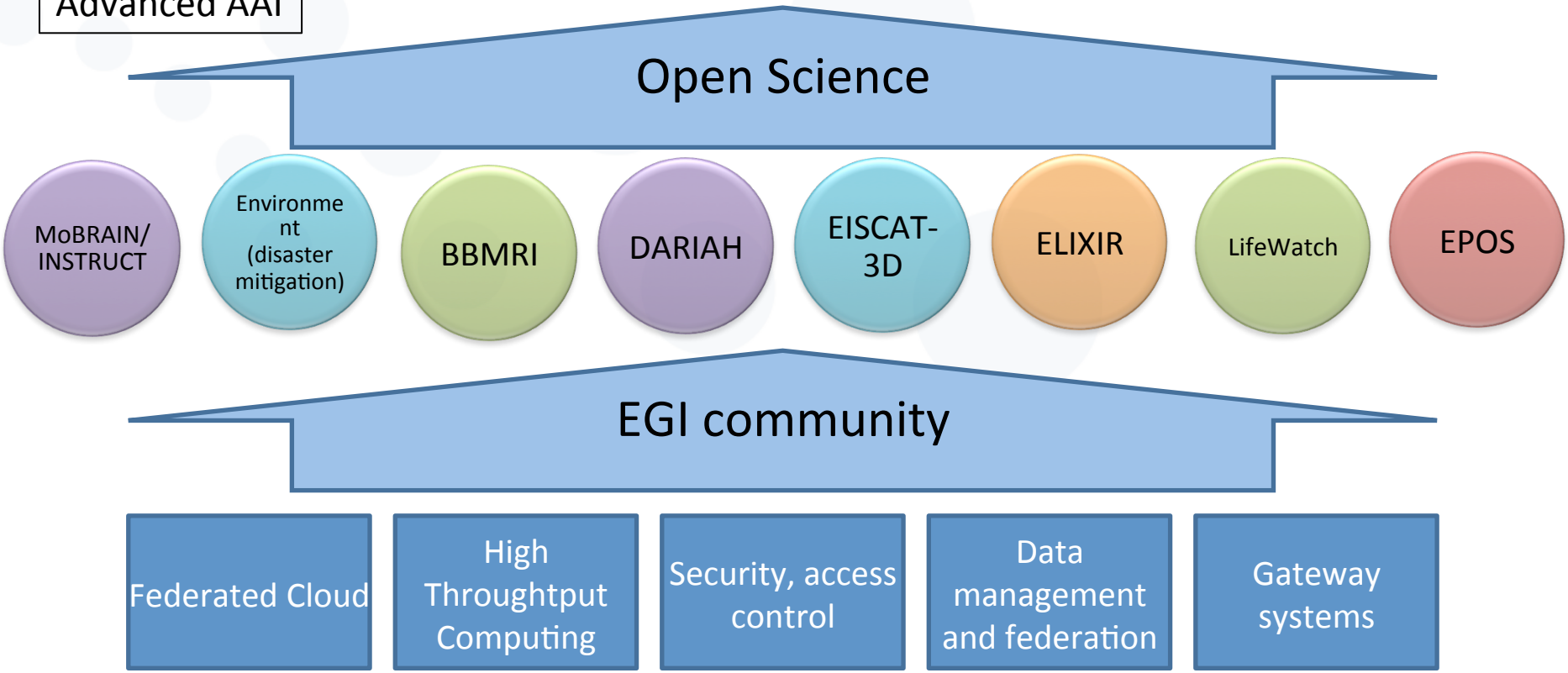


# Outcomes

Community clouds | Data products | Active Repositories | Training resources

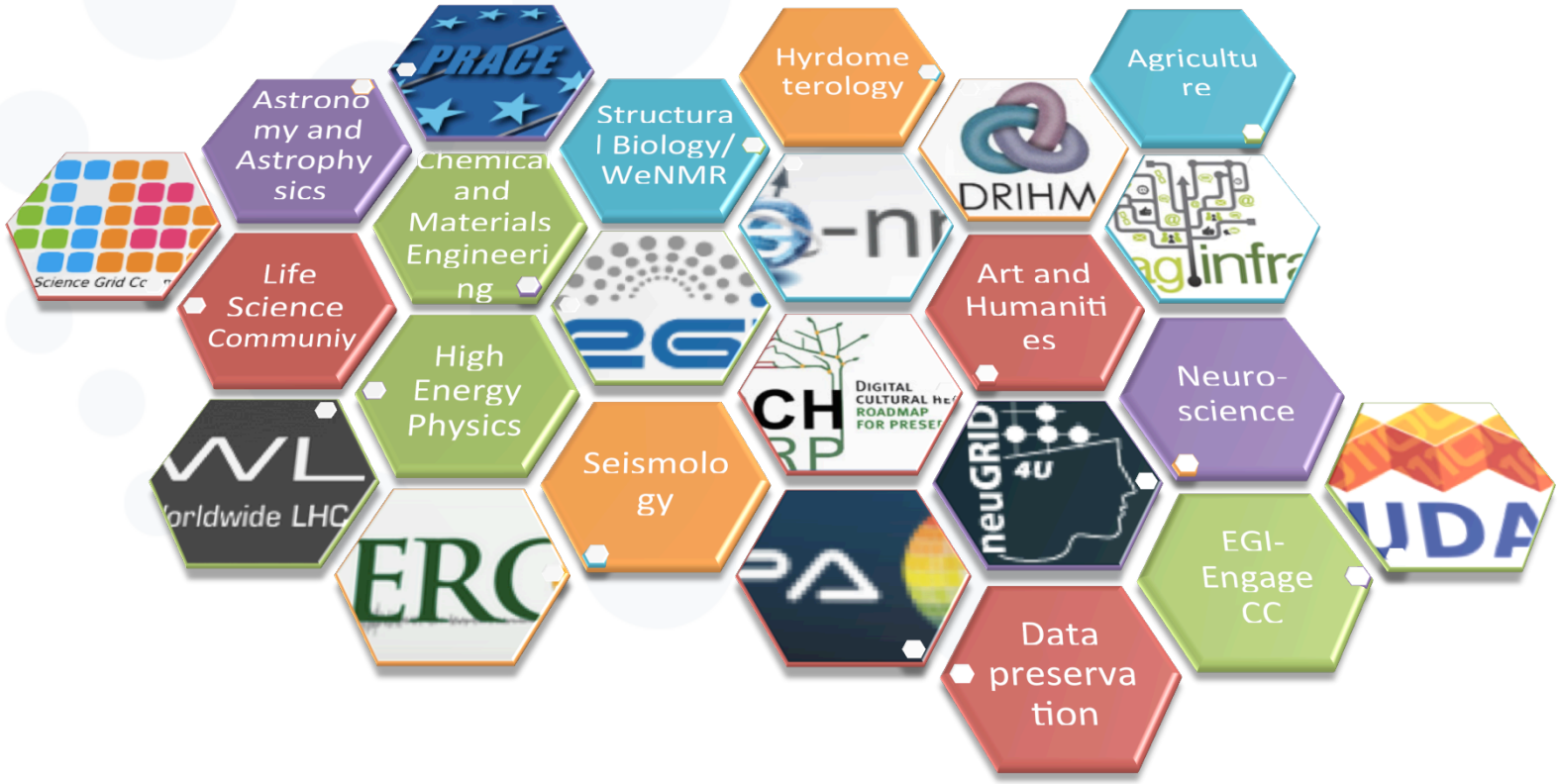
GPU federations | Science gateways | Scalable applications | User support

Advanced AAI



# Actors – present and future

# Federate Knowledge in Europe




Join the Competence Centre meetings, every day 17:00 – 18:00, OPEN!

# The big shifts

New governance to community engagement

The Distributed Competence Centre

Better services for the long tail

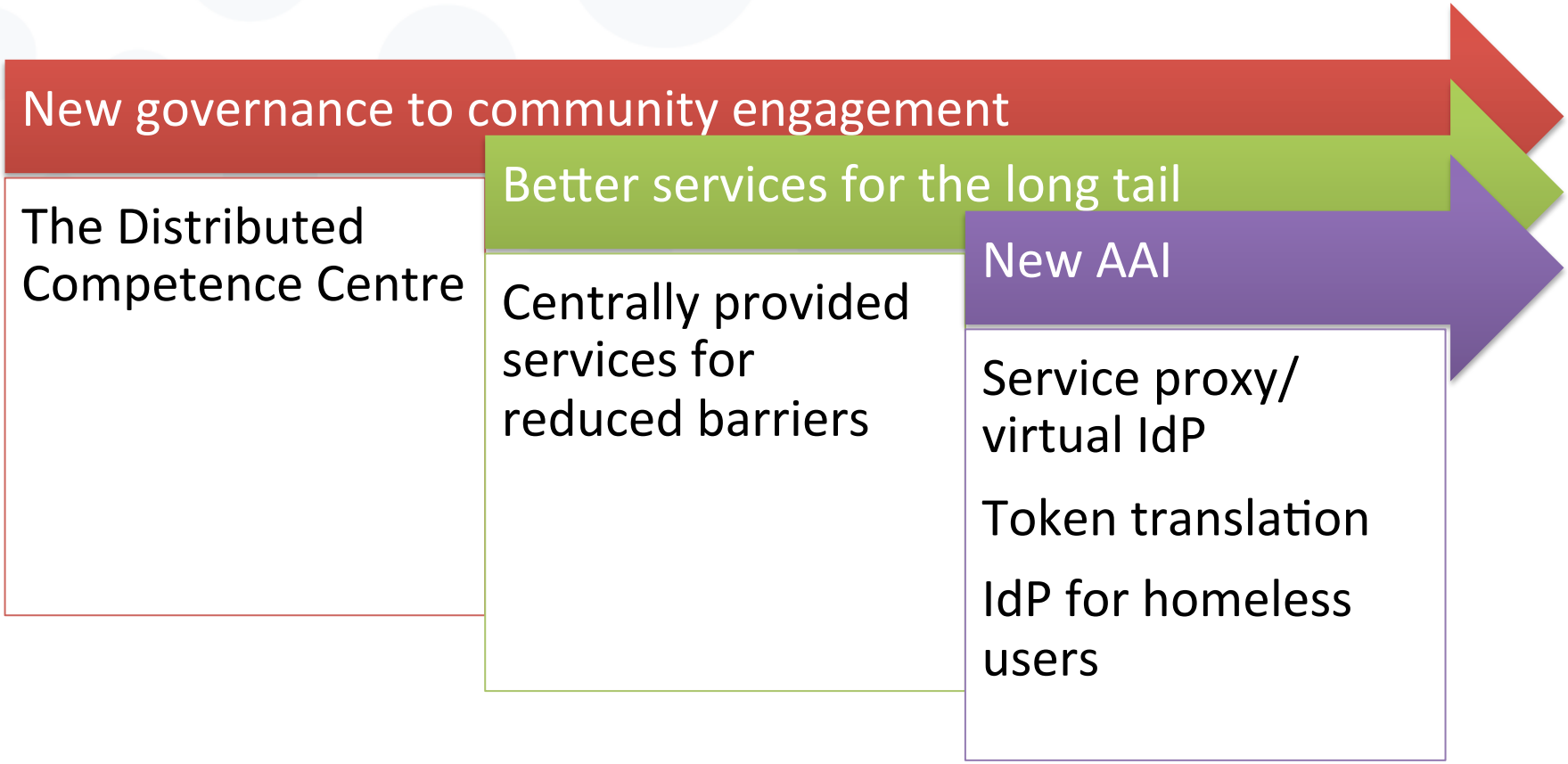
Centrally provided services for reduced access barriers

# Services for the long tail of science

- Move towards a “zero (technical) barrier” e-infrastructure
  - Services dedicated to individual users or very small collaborations:
    - No certificate, no VO, full EGI experience
- User facing features
  - Log in using their federated identity
  - Provide the additional information not available in the IdP
  - Discover (marketplace) and submit a request for resources
- EGI/NGIs facing features:
  - Assign UIDs to users of the long tail of science platform
  - Approve user request
  - Monitor usage of resources

Sessions on Wed 20/5

# The big shifts

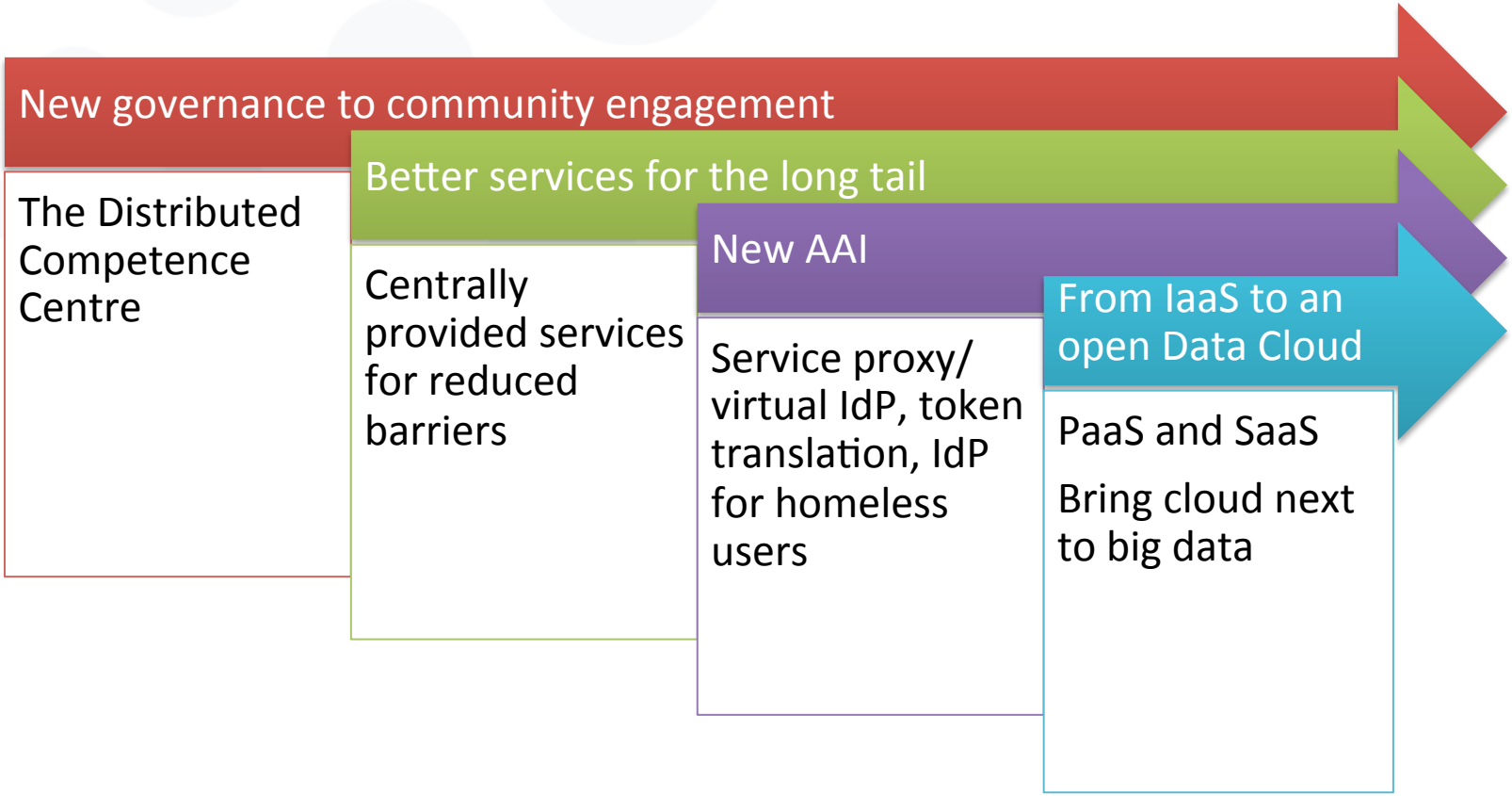


# Advancing AAI

- EGI users are directly/indirectly using x509 credentials to access the production services
- Objective: allow users to use their existing institutional credentials by
  - Replicating the current architecture to manage user communities in the other authentication technologies already used by the users
  - Integrating other federated identities into EGI services
- Testing and deployment of AAI services, and requirements analysis in close collaboration with the CCs and the other communities
  - Catch all IdP service (EGI sso), online CA, attribute authorities to manage users without X.509 ceritificate
  - Service proxy/Virtual IdP: technical service AND support to help communities to integrate easily their IdP with EGI. Integrating new IdP and attribute authorities in a one-step.
- Collaboration with AARC project

AAI track on Friday 22/05

# The big shifts

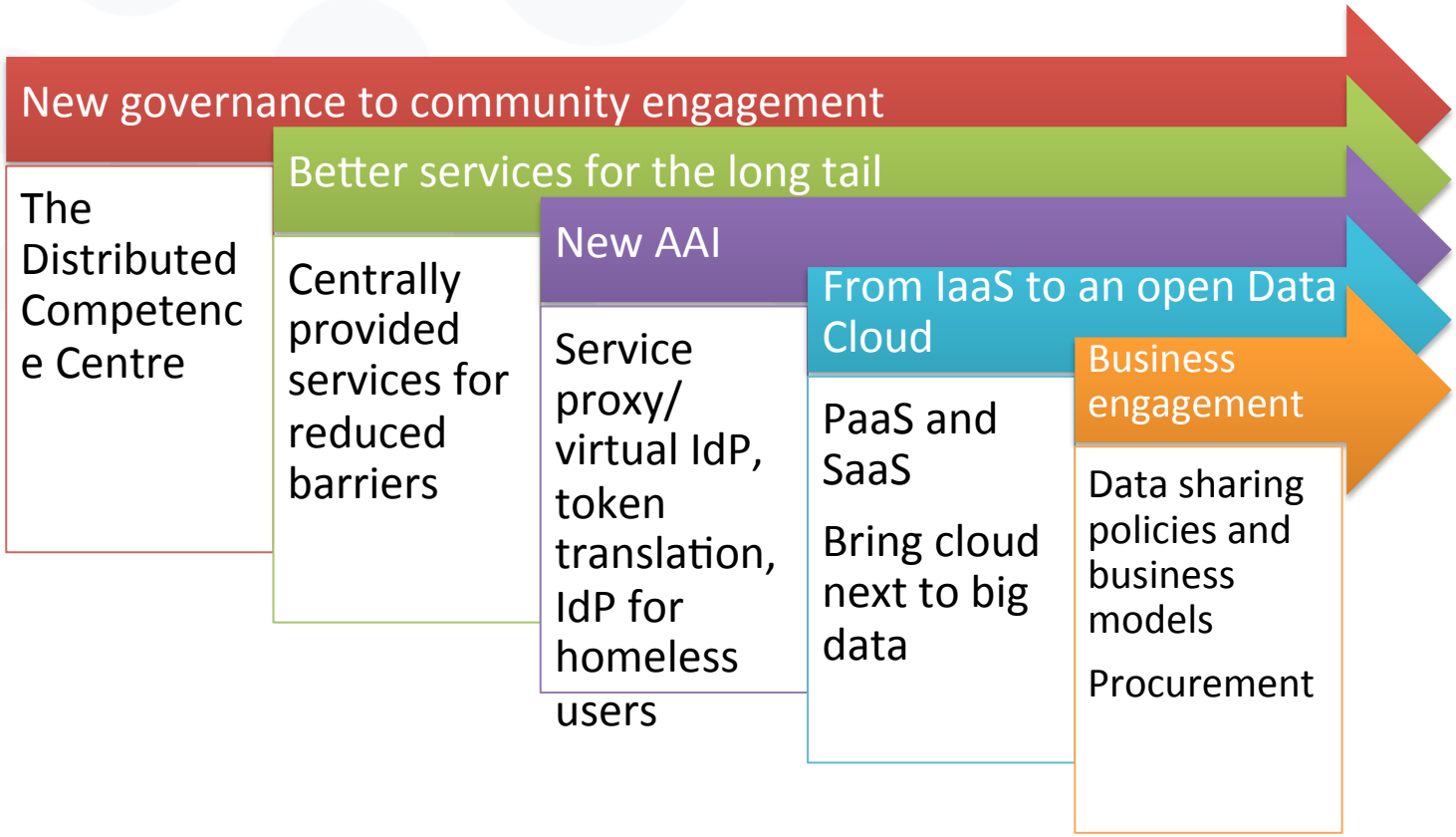


# Federated Cloud + Open Data: Open Data Cloud

- Objective: scalable access to open research data for discovery, access and use
- Remove policy and technical barriers
  - Bring cloud service next to distributed data repositories
  - Replicate open research data of research/commercial relevance
  - Discovery, accounting
  - Provide PaaS and SaaS and evolve the federation services
  - Virtual appliance library of community tools and data for
    - Repeatability of science, training and education (EDISON)
- Collaboration with EUDAT and INDICO-Datacloud
- Multiple stakeholders involved

Open Data Cloud track on Thursday 21/05

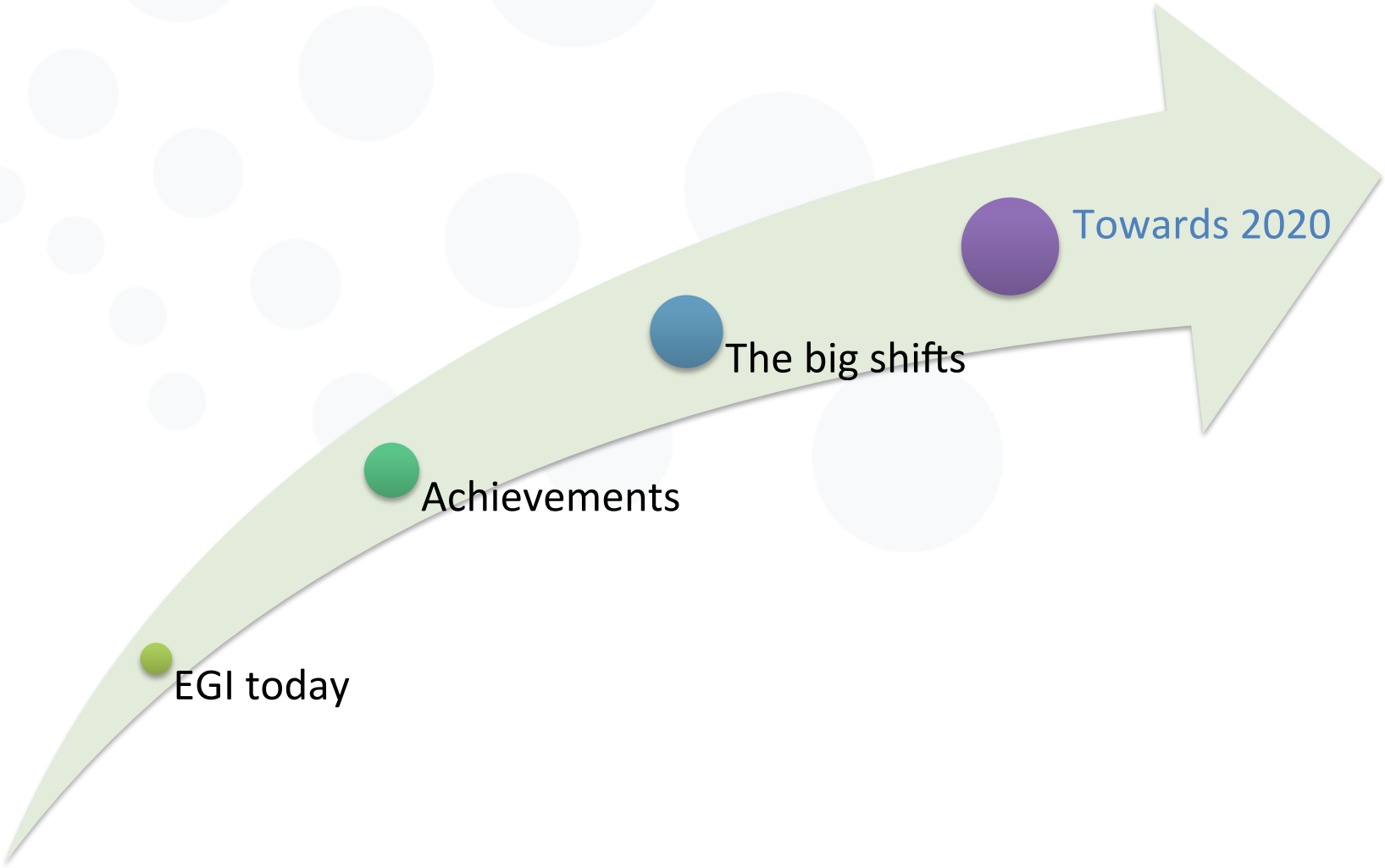
# The big shifts



# Policy and business

- Pay-for-use and cross-border procurement
- Facilitate collaboration with SMEs (focus on consumer side) via a model to be adopted and adapted for a wider number of NGIs/Resource Centres
  - Use cases from agriculture, fishery and marine sciences, biodiversity, earth science
- Explore with SMEs opportunities and threats around the Open Data and co-develop business models for their exploitation
    - Market analysis and user requirements
    - Data Sharing Policies and Legal Aspects

Sessions on Wed, Thu and Fri

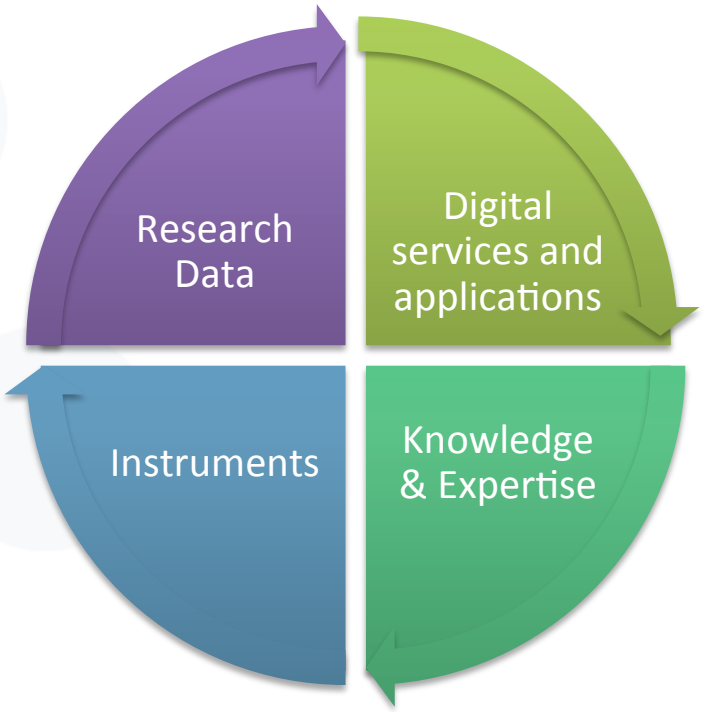

Towards 2020
The big shifts
Achievements
EGI today

# Digital ERA – State of play 2015

- Incomplete national roadmaps for Research and e-Infrastructures
  - E-Infrastructures and RIs should be components of the same research system
- e-Infrastructure Commons not fully achieved yet
  - Lack of e-Infrastructure capacity for multidisciplinary research and the long tail of science
  - Different access policies for user groups in each access
  - Incomplete technical interoperability, different access policies
  - The "Commons" governance principle not widely adopted
  - Non organized landscape of multiple service providers and research communities, lack of cross-border procurement/funding scheme that allows coordinated resource management across Europe (except for GEANT)
- Lack of one 'backbone' of European ICT capabilities

# Open Science a Complex Resource System

- Shared resources
  - Integrated, easy and fair access
- Engaged communities
  - Participating in the process
  - Culture of sharing
  - Collaborating in the management and stewardship
- Governance
  - Rules to access
  - Rules to resolve conflicts
  - Rules to balance quality vs. openness
- Financial support
  - For long-term availability

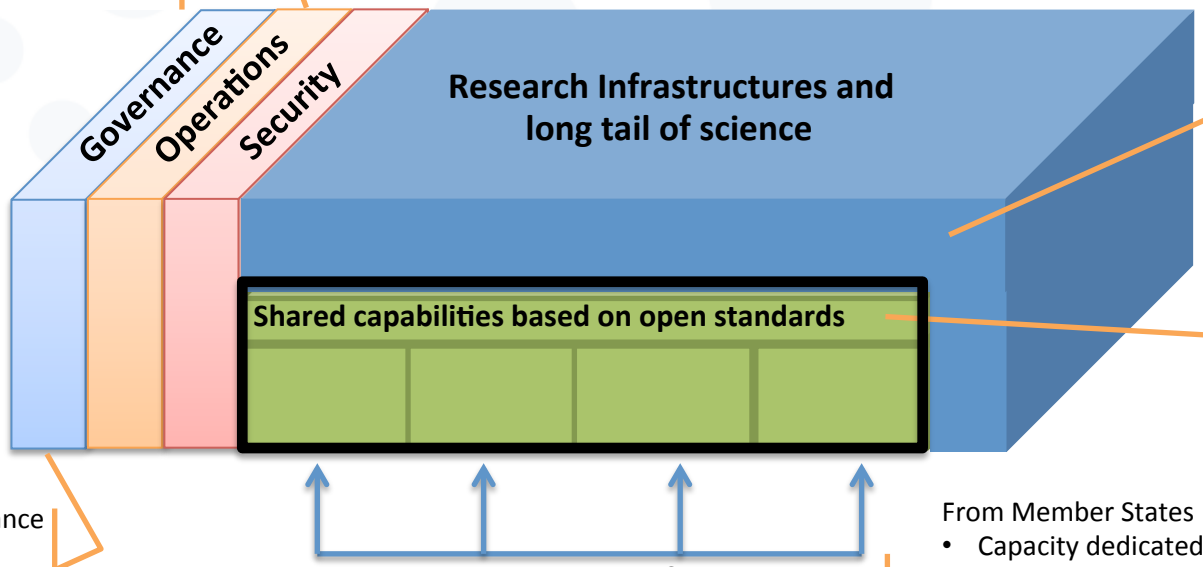


# A common endeavor (EU perspective)

# Developing an OSC: Shared Open Science Infrastructure Backbone

Federated operations and support
- Service desk
- Monitoring and accounting
- Capacity management
- Service level management

- Network of CSIRT
- Federated IdPs, Auth and Authz
- Management of different levels of assurance




- Research platform built on top of shared capabilities plus community owned resources
- Data products, tools, scientific gateways, virtual labs

Core capabilities
- Open Science Cloud (e.g., VM management, Data storage/access/ discovery)
- PID
- Service registry and marketplace

Multi-level governance with community participation
- Local
- National
- European

From Member States
- Capacity dedicated to large RIs
- Free pools for long tail researchers
- Both pubiicly funded and commercial providers (all supporting open standards and no lock-in)

# How can EGI contribute?

Federate digital capabilities, resources and expertise

Operate services across the federated infrastructure

Co-create and integrate open and user-driven services and solutions

Be a trusted adviser on data and compute intensive science

# EGI Vision

Researchers from all disciplines
have easy, integrated and open access
to the advanced digital capabilities,
resources and expertise
needed to collaborate and to carry out
compute/data intensive science and
innovation

# EGI Mission

Create and deliver open solutions
for science and research
infrastructures
by federating digital capabilities,
resources and expertise
across communities and national
boundaries

# Thank you for your attention.

*Questions?*

www.egi.eu